

Which Is Better, an Ensemble of Positive–Negative Pairs or a Centered Spherical Simplex Ensemble?

XUGUANG WANG

The Pennsylvania State University, University Park, Pennsylvania

CRAIG H. BISHOP

UCAR, Boulder, Colorado, and Naval Research Laboratory, Monterey, California

SIMON J. JULIER

Falls Church, Alexandria, Virginia

(Manuscript received 30 July 2003, in final form 8 January 2004)

ABSTRACT

New methods to center the initial ensemble perturbations on the analysis are introduced and compared with the commonly used centering method of positive–negative paired perturbations. In the new method, one linearly dependent perturbation is added to a set of linearly independent initial perturbations to ensure that the sum of the new initial perturbations equals zero; the covariance calculated from the new initial perturbations is equal to the analysis error covariance estimated by the independent initial perturbations, and all of the new initial perturbations are equally likely. The new method is illustrated by applying it to the ensemble transform Kalman filter (ETKF) ensemble forecast scheme, and the resulting ensemble is called the spherical simplex ETKF ensemble. It is shown from a multidimensional Taylor expansion that the symmetric positive–negative paired centering would yield a more accurate forecast ensemble mean and covariance than the spherical simplex centering if the ensemble were large enough to span all initial uncertain directions and thus the analysis error covariance was modeled precisely. However, when the number of uncertain directions is larger than the ensemble size, the spherical simplex centering has the advantage of allowing almost twice as many uncertain directions to be spanned as the symmetric positive–negative paired centering. The performances of the spherical simplex ETKF and symmetric positive–negative paired ETKF ensembles are compared by using the Community Climate Model Version 3 (CCM3). Each ensemble contains 1 control forecast and 16 perturbed forecasts. The NCEP–NCAR reanalysis data for the boreal summer in 2000 are used for the initialization of the control forecast and the verifications of the ensemble forecasts. The accuracy of the ensemble means, the accuracy of predictions of forecast error variance, and the ability of the ETKF ensembles to resolve inhomogeneities in the observation distribution were all tested. In all of these test categories, the spherical simplex ETKF ensemble was found to be superior to the symmetric positive–negative paired ETKF ensemble. The computational expense for generating spherical simplex ETKF initial perturbations is about as small as that for the symmetric positive–negative paired ETKF. Also shown is that the seemingly straightforward centering method, in which centered perturbations are obtained by subtracting the average of the perturbations from each individual perturbation, is unsatisfactory because the covariance estimated by the uncentered perturbations is not necessarily conserved after centering.

1. Introduction

The ensemble transform Kalman filter (ETKF) ensemble generation technique introduced in Wang and Bishop (2003) did not give an explicit method whereby the initial ensemble perturbations could be centered about the best available estimate of the true state, that is, the analysis. This is undesirable because, ideally, one would like the ensemble mean to *always* be equal to

the minimum error variance estimate of the true state. Our primary aim in this paper is to compare the effect of differing ensemble centering procedures on the performance of the ETKF ensemble.

The question of how one should center an ensemble does not appear to have received much attention in published literature. Operationally, the singular vector (SV) scheme at the European Centre for Medium-Range Weather Forecasts (ECMWF; Buizza and Palmer 1995; Molteni et al. 1996) and the breeding scheme at the National Centers for Environmental Prediction (NCEP; Toth and Kalnay 1993, 1997) both select symmetric positive–negative paired centering, in which initial per-

Corresponding author address: Xuguang Wang, Dept. of Meteorology, The Pennsylvania State University, 503 Walker Building, University Park, PA 16802.
E-mail: xuguang@essc.psu.edu

turbations are created by letting half of these perturbations be the negative of the other half. Toth and Kalnay (1997) also proposed another method in which centered perturbations were obtained by removing the average of the perturbations from each individual perturbation vector. Hereafter we call it the subtract-mean method. Toth and Kalnay (1997) found that the breeding ensemble centered by the subtract-mean method was less skillful than that centered by the symmetric positive–negative paired method, and they also mentioned that the ECMWF found the same results for their SV scheme. We revisited this subtract-mean method for the breeding scheme, and the results are discussed in section 5.

Beside making the sum of the perturbations equal to zero, there are two additional aspects of the ensemble that one needs to control when centering the initial perturbations. First, if the ensemble covariance is going to be used to estimate the forecast error covariance then one would like the analysis error covariance estimated by the initial perturbations to be preserved before and after centering; second, if ensemble forecast perturbations are to be treated as equally likely error realizations, then the centered analysis perturbations must also be equally likely.

The traditional symmetric positive–negative pair method satisfies the above requirements. In this paper we describe a new centering method, called the spherical simplex method. As shown in appendix A, if the ensemble size were large enough to span *all* uncertain directions and the analysis error covariance could be modeled perfectly by the outer product of the initial ensemble perturbations, symmetric positive–negative paired centering provides third-order-accurate ensemble mean and ensemble covariance. Hence it is superior to the spherical simplex centering that only provides second-order accuracy. But if the ensemble size is not sufficiently large, which is true for a system with high dimensions such as a typical atmospheric forecast model, the spherical simplex centering has the advantage of allowing almost twice as many uncertain directions to be spanned as the symmetric positive–negative paired centering (see appendix A). When the ensemble size is less than the number of uncertain directions, there is no readily available theoretical basis for determining the extent to which one centering scheme would outperform the other. Consequently, we empirically tried both schemes in order to determine which centering approach yielded the most useful ensemble. Also note that for computationally inexpensive ensemble generation schemes such as the breeding and the ETKF schemes, there is no apparent computational advantage in using the common symmetric positive–negative paired centering relative to the spherical simplex centering. The goal of this paper is to test both centering methods on the ETKF scheme and to answer the question, which is better, the symmetric positive–negative paired ETKF ensemble or the spherical simplex ETKF ensemble?

In section 2 we introduce the theory of the spherical simplex ETKF and the symmetric positive–negative paired ETKF ensemble generation schemes. We also demonstrate that the subtract-mean centering is sub-optimal since the ETKF estimated analysis error covariance is not necessarily preserved. Section 3 describes briefly how the numerical experiment is designed. Section 4 compares the performance of the spherical simplex ETKF and symmetric positive–negative paired ETKF in terms of short-term ensemble subspace rank, the skills in estimating analysis error variance, and the accuracy of the ensemble mean and ensemble variance. In section 5 we summarize and discuss the results.

2. Theory of centering ETKF initial perturbations

a. Review of one-sided ETKF initial perturbations

The ETKF analysis perturbations are achieved by postmultiplying the short-term ensemble forecast perturbations by a transformation matrix. This transformation matrix is obtained by solving the error covariance update equation for an optimal data assimilation scheme within the ensemble subspace [please see Bishop et al. (2001); Wang and Bishop (2003) for the ETKF theory]. One-sided ETKF ensemble refers to the ETKF ensemble whose initial perturbations are not explicitly centered about the analysis. Specifically in our current experiment, we define K forecast perturbations at the 12-h forecast lead time as

$$\mathbf{X}^f = (\mathbf{x}_1^f - \bar{\mathbf{x}}^f, \mathbf{x}_2^f - \bar{\mathbf{x}}^f, \dots, \mathbf{x}_K^f - \bar{\mathbf{x}}^f), \quad (1)$$

where \mathbf{x}_i^f , $i = 1, \dots, K$, are K 12-h forecasts¹ and $\bar{\mathbf{x}}$ is the mean of the K 12-h forecasts, that is,

$$\bar{\mathbf{x}} = (\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_K)/K. \quad (2)$$

The ensemble-estimated 12-h ensemble mean forecast error covariance matrix is

$$\mathbf{P}^f = \mathbf{Z}^f(\mathbf{Z}^f)^T, \quad (3)$$

where $\mathbf{Z}^f = \mathbf{X}^f/\sqrt{K}$.² The analysis perturbations \mathbf{X}^a are obtained by postmultiplying (1) by a transformation matrix \mathbf{T} , that is,

$$\mathbf{X}^a = \mathbf{X}^f\mathbf{T}. \quad (4)$$

¹ Note, as discussed in section 3 and also for the reason to maintain fixed ensemble size during ensemble forecast cycling, in our experiment below the K 12-h forecasts only include the K perturbed 12-h forecasts without the 12-h control run.

² Note that in Eqs. (A2)–(A7) of appendix A the weights used to calculate the mean and the covariance are the same, specifically in both (2) and (3), $w_i = 1/K$ for all i [see (A2)–(A7) for w_i]. Division by K in (3) provides the maximum likelihood estimate of the covariance (Wilks 2002), whereas division by $K - 1$ provides the corresponding unbiased estimate (Ross 1998). Both choices are reasonable. To be consistent with the derivation in appendix A, we choose the division by K in (3) and in all other variance and covariance calculations.

The transformation matrix \mathbf{T} is derived from the error covariance update equation for an optimal data assimilation scheme provided that forecast error covariance is given by Eq. (3) [see Bishop et al. (2001) and Wang and Bishop (2003) for the derivation]. The result is $\mathbf{T} = \mathbf{C}(\mathbf{\Gamma} + \mathbf{I})^{-1/2}$, where \mathbf{C} and $\mathbf{\Gamma}$ are the eigenvector and eigenvalue matrices of $(\mathbf{X}^f)^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{X}^f / K$, in which \mathbf{H} is the observation operator and \mathbf{R} is the observation error covariance matrix.

Note only $K - 1$ independent ETKF analysis perturbations are generated from (1)–(4). This is because the sum of the K forecast perturbations in (1) is zero and therefore the last (i.e., the smallest) eigenvalue of $\mathbf{\Gamma}$ is equal to zero (Note that throughout this paper, the eigenvectors and corresponding eigenvalues are organized from left to right in order of decreasing eigenvalues). Thus, (1) postmultiplied by the last column of \mathbf{C} is a zero vector. In other words, the K th analysis perturbation is a zero vector. The $K - 1$ nonzero ETKF analysis perturbations can then be written as

$$\mathbf{X}^a = (\mathbf{x}_1^a, \mathbf{x}_2^a, \dots, \mathbf{x}_{K-1}^a) = \mathbf{X}^f \mathbf{\Theta} (\mathbf{F} + \mathbf{I})^{-1/2}, \quad (5)$$

where $\mathbf{\Theta}$, a $K \times (K - 1)$ matrix, contains the first $K - 1$ columns of \mathbf{C} , and \mathbf{F} is a $(K - 1) \times (K - 1)$ diagonal matrix whose diagonal elements contain the first $K - 1$ eigenvalues in $\mathbf{\Gamma}$. The sum of columns in \mathbf{X}^a is not zero as they are orthogonal to each other in the observation space [Eq. (23) in Wang and Bishop 2003]. In other words, the ETKF analysis perturbations given by (5) are not centered about the analysis. As discussed in the introduction and appendix A, this is not desirable for ensemble forecasting where the ensemble mean is used to provide the minimum error variance estimate of the true state and the ensemble covariance is used to estimate the corresponding error covariance of this estimate. The goal of this work is to test different centering schemes on the ETKF analysis perturbations.

b. Constraints on centering

To center the ETKF initial perturbations, beside making the sum of the perturbations equal to zero, there are two more aspects of the one-sided ETKF analysis perturbations in (5) that we need to preserve. First, the analysis error covariance estimated by the one-sided ETKF [see Eq. (18b) in Bishop et al. 2001] is

$$\mathbf{P}^a = \mathbf{Z}^f \mathbf{\Theta} (\mathbf{F} + \mathbf{I})^{-1} \mathbf{\Theta}^T (\mathbf{Z}^f)^T = \mathbf{X}^a (\mathbf{X}^a)^T / K. \quad (6)$$

Note the ETKF-estimated analysis error covariance given by (6) would be precisely equal to the true analysis error covariance given that the true forecast error variance were equal to (3) and the data assimilation scheme were optimal (Bishop et al. 2001; Wang and Bishop 2003). In other words, the one-sided ETKF analysis perturbations are the exact square roots of the analysis error covariance (Tippett et al. 2003). Therefore, when centering the one-sided ETKF initial perturbations, the analysis error covariance given by (6) needs to be con-

served. Note this constraint tries to maintain the second-order-accurate ensemble mean and covariance (appendix A).

Second, as shown in appendix B the one-sided ETKF initial perturbations projected onto observation space and normalized by the root-mean-square (rms) observation error are equally likely. For the end users, the ensemble outputs are easy to interpret if the ensemble members are designed to be equally likely. For example, evaluation of the ensemble spread by the rank histogram (Hamill 2001) automatically assumes that each ensemble member is equally likely. If the amplitude of one initial perturbation were improbably larger than the others, we would have to assign different weight to this member when calculating ensemble mean and ensemble covariance. To avoid such complications, we also require that the centered initial ETKF perturbations maintain the characteristic of being equally likely. This constraint will affect the higher-order accuracy of the ensemble mean and ensemble covariance (appendix A).

c. Spherical simplex centering for the ETKF

To achieve spherical simplex centering, one post-multiplies \mathbf{X}^a in (5) by a $(K - 1) \times K$ matrix \mathbf{U} to form K perturbations

$$\mathbf{Y}^a = \mathbf{X}^a \mathbf{U} = (\mathbf{y}_1^a, \mathbf{y}_2^a, \dots, \mathbf{y}_K^a). \quad (7)$$

Note hereafter the initial state ensemble is defined to include the analysis and K perturbed members. The matrix \mathbf{U} is selected to ensure that (a) the sum of \mathbf{y}_i^a , $i = 1, \dots, K$ is zero; (b) the analysis error covariance estimated by the outer product of (7) is equal to (6); and (c) the centered perturbations projected onto observation space and normalized by the root-mean-square observation error, denoted as $\tilde{\mathbf{H}} \mathbf{y}_i^a$, $i = 1, \dots, K$, are equally likely (see appendix B for the meaning of “equally likely”).

A sufficient condition to meet requirement (a) is that

$$\mathbf{U} \mathbf{1} = \mathbf{0}, \quad (8)$$

where $\mathbf{0}$ is a vector with each element equal to zero, and $\mathbf{1}$ is a vector with each element equal to one. From (5), (6), and (7), the requirement (b) of preserving the ETKF-estimated analysis error covariance after centering is satisfied whenever

$$\mathbf{U} \mathbf{U}^T = \mathbf{I}, \quad (9)$$

where \mathbf{I} is the identity matrix. To satisfy the requirement (c), that is, to let the value of the probability density function (pdf) of $\tilde{\mathbf{H}} \mathbf{y}_i^a$ be equal to each other for all i , first note that from requirement (b) the analysis error covariance in observation space estimated by $\tilde{\mathbf{H}} \mathbf{Y}^a$ is equal to (B4). Also, from (B3) and (7),

$$\tilde{\mathbf{H}} \mathbf{Y}^a = \tilde{\mathbf{H}} (\mathbf{y}_1^a, \mathbf{y}_2^a, \dots, \mathbf{y}_K^a) = \sqrt{K} \mathbf{E} \mathbf{D}^{1/2} \mathbf{U}. \quad (10)$$

Substituting (B4) and (10) into (B1) to replace \mathbf{P} and \mathbf{x}' , we find that in order to satisfy the requirement (c)

the diagonal elements of $\mathbf{U}^T\mathbf{U}$ must be equal to each other, that is, each column of \mathbf{U} must have the same magnitude.

There is more than one \mathbf{U} satisfying the three requirements. So far we have found two easy solutions, shown in appendix C. Because the matrix \mathbf{U} comes from the concepts of spherical simplex sigma points [see appendixes A and C and/or Julier and Uhlmann (2003)], we call the ETKF analysis perturbations constructed this way the spherical simplex ETKF. Note also that the third requirement for designing the simplex points depends on user's interest. For example, Julier and Uhlmann (2002a) chose the simplex points and the weights in a way that would minimize the skewness of the simplex points so that the errors in the estimate of the mean and covariance associated with the third-order moment—that is, the skewness—were minimized.

Because the ensemble size is much smaller than the number of directions to which the true error variance projects, the data assimilation scheme in reality is not optimal and the model is not perfect, so the error variance is significantly underestimated by the ETKF. The inflation factor method introduced in Wang and Bishop (2003) is used to ameliorate this problem. The idea is to multiply the initial perturbations in (7) by an inflation factor to ensure that 12-h ensemble forecast variance is consistent with the 12-h control forecast error variance over global observation sites [please see section 2c of Wang and Bishop (2003) for details]. At each perturbation update time the maximum likelihood parameter estimation theory (Dee 1995) is used to check this consistency and calculate an instantaneous inflation factor. The overall inflation factor used to inflate the perturbations in (7) is the product of all previous and current instantaneous inflation factors. In our experiment, the instantaneous inflation factor converges to 1, and the overall inflation factor oscillates about a constant within 1 week of 12-h perturbation initialization and forecast cycles. This quick convergence results from the fact that for the global observational network, the number of independent elements in the innovation vector is large. In this way, the value of the overall inflation factor is automatically determined by the ensemble forecasting system itself.

d. Symmetric positive–negative paired centering for the ETKF

In the traditional symmetric positive–negative paired centering scheme, an ensemble of K initial perturbations is created by letting half of these perturbations be the negative of the other half. In this section, we apply the symmetric positive–negative paired centering scheme to the ETKF initial perturbations. Specifically, K positive–negative paired ETKF initial perturbations are built from $K/2$ independent one-sided ETKF perturbations. From K 12-h ensemble perturbations defined in the same way

as in Eq. (1), optimal truncation³ is used to obtain $K/2$ one-sided ETKF perturbations, that is,

$$\begin{aligned}\mathbf{X}^a &= \mathbf{X}^f \mathbf{T}_t = \mathbf{X}^f \mathbf{C}_t (\mathbf{\Gamma}_t + \mathbf{I}_t)^{-1/2} \\ &= (\mathbf{x}_1^{a'}, \mathbf{x}_2^{a'}, \dots, \mathbf{x}_{K/2}^{a'}),\end{aligned}\quad (11)$$

where the subscript t denotes the truncation. The diagonal elements of the $K/2 \times K/2$ matrix $\mathbf{\Gamma}_t$ contains the largest $K/2$ eigenvalues of $(\mathbf{X}^f)^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{X}^f / K$. The matrix \mathbf{C}_t contains the corresponding $K/2$ eigenvectors. Because the variance of the 12-h forecast perturbations starting from the positive–negative paired analysis perturbations are mainly distributed in the first half eigenvectors (see also Fig. 2 later), almost all variance is maintained in the truncated space associated with the one-sided perturbations in (11). The $K/2$ pairs of perturbations \mathbf{Y}^a are then built up as

$$\mathbf{Y}^a = \frac{1}{\sqrt{2}} \cdot (\mathbf{x}_1^{a'}, \mathbf{x}_2^{a'}, \dots, \mathbf{x}_{K/2}^{a'}, -\mathbf{x}_1^{a'}, -\mathbf{x}_2^{a'}, \dots, -\mathbf{x}_{K/2}^{a'}), \quad (12)$$

where the coefficient $1/\sqrt{2}$ is selected to conserve the variance. Mathematically, Eq. (12) is equivalent to postmultiplying \mathbf{X}^a in (11) by a $K/2 \times K$ matrix \mathbf{S} , that is,

$$\mathbf{S} = \frac{1}{\sqrt{2}} \{\mathbf{I}(-\mathbf{I})\}, \quad (13)$$

where $\{\mathbf{I}(-\mathbf{I})\}$ is a $K/2 \times K$ matrix. The $K/2 \times K/2$ identity matrix \mathbf{I} constructs the first $K/2$ columns and $-\mathbf{I}$ constructs the remaining $K/2$ columns. It is easy to verify that \mathbf{S} also satisfies requirements (a), (b), and (c) in section 2c. However, it only contains $K/2$ independent directions while the \mathbf{U} matrix in (7) has $K - 1$ independent directions. The inflation factor method of Bishop and Wang (2003) is also applied to the symmetric positive–negative paired ETKF ensemble.

e. Problem in subtract-mean centering

In the subtract-mean method, centered perturbations are obtained by removing the mean of the perturbations from each individual perturbation vector. Such an approach would be appropriate if one had a set of perturbations that represented a random sample of analysis errors. However, square root filters such as the ETKF described in Bishop et al. (2001) provide analysis perturbations that are not, strictly speaking, random samples drawn from the distribution of the analysis error. Instead, as discussed in section 2a, they are the square roots of the analysis error covariance matrix. Thus, as shown in Eq. (6), the analysis error covariance is given by the outer product of these one-sided initial perturbations without removing the mean of the one-sided initial perturbations. As discussed below, we do not

³ We call the truncation in (11) “optimal” as the eigenvectors corresponding to the first-half *largest* eigenvalues are maintained.

want to apply the subtract-mean centering to the ETKF because the ETKF-estimated analysis error covariance given by (6) will not be preserved after the mean is subtracted. Specifically, suppose we have K one-sided ETKF initial perturbations contained in K columns of \mathbf{X}^a and the analysis error covariance is estimated by its outer product as in (6). If written in the similar format of (7), the subtract-mean centering is equivalent to post-multiplying \mathbf{X}^a by a $K \times K$ matrix \mathbf{V} , where

$$\mathbf{V} = \begin{pmatrix} 1 - 1/K & -1/K & \cdots & -1/K \\ -1/K & 1 - 1/K & \cdots & -1/K \\ \vdots & \vdots & \ddots & \vdots \\ -1/K & -1/K & \cdots & 1 - 1/K \end{pmatrix}. \quad (14)$$

Although \mathbf{V} satisfies requirement (a) and (c) in section 2c, it is easy to verify that requirement (b) is not guaranteed to be satisfied because

$$\mathbf{V}\mathbf{V}^T = \mathbf{V} \neq \mathbf{I}. \quad (15)$$

In fact, both the initial ensemble variance and the ensemble subspace rank are changed after the subtract-mean centering. This can be further illustrated by Fig. 1. In Fig. 1a, assume two one-sided ETKF initial perturbations locate at (0, 1) and (1, 0). After subtracting the average of these two one-sided perturbations, the new perturbations locate in Fig. 1b at (1/2, -1/2) and (-1/2, 1/2). So, after subtract-mean centering the two-dimensional one-sided ETKF ensemble subspace collapses to one dimension and the variance is reduced to half of that of the one-sided ETKF perturbations. Note the ETKF-estimated analysis error covariance is given by the outer product of (0, 1) and (1, 0). So after subtract-mean centering, the outer product of (1/2, -1/2) and (-1/2, 1/2) does not preserve the ETKF-estimated analysis error covariance. Since subtract-mean centering cannot satisfy our three constraints, we did not apply it to the ETKF ensemble. However, as mentioned in section 5, we tried the subtract-mean centering for the breeding ensembles following Toth and Kalnay (1997).

3. Numerical experiment design

In our experiment, the ensemble includes 1 control forecast and 16 perturbed forecasts, that is, $K = 16$. We used the same numerical model CCM3 (Jeffery et al. 1996) at T42 resolution as in Wang and Bishop (2003). We also used the NCEP-NCAR reanalysis (Kalnay et al. 1996) as the analysis and verification. The time period we consider is the Northern Hemisphere summer in year 2000. The observational network was also assumed to contain only rawinsonde observations. Pseudo-observations were obtained from the reanalysis data by relabeling reanalysis values of wind and temperature at the rawinsonde sites as “observations.” The observation error covariance matrix was assumed to be time independent and diagonal. To estimate the error variance of these pseudo-observations, we first calculate 12-h

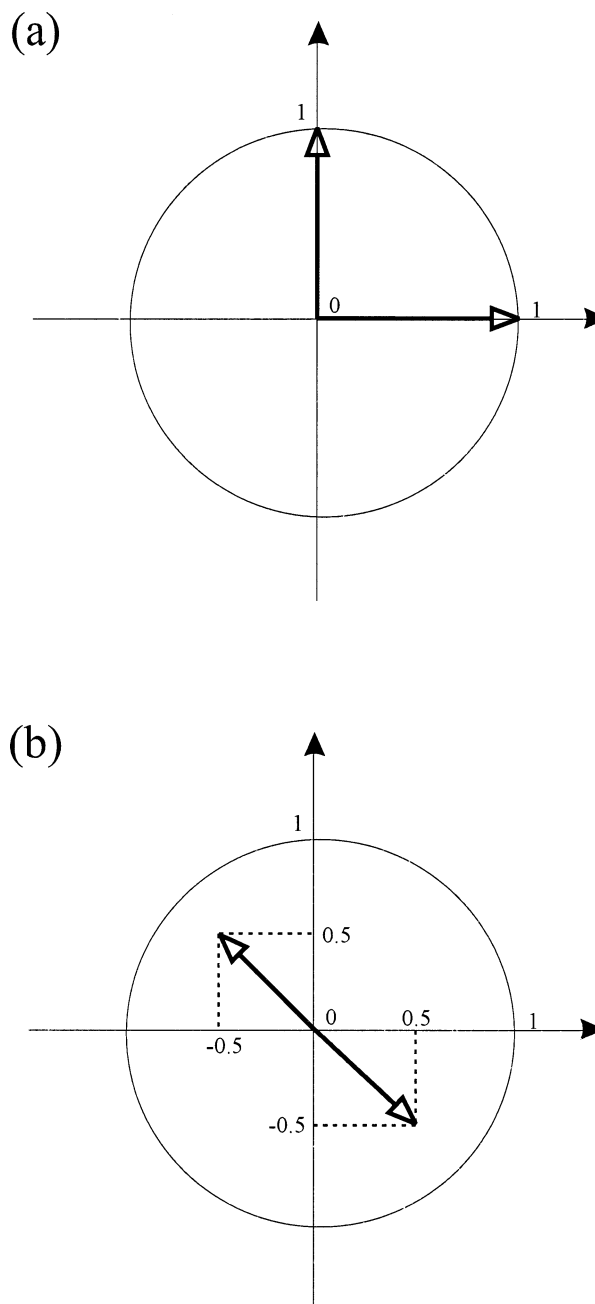


FIG. 1. Illustration that the subtract-mean centering scheme does not preserve the covariance. (a) Assume two one-sided 2D perturbations at (0, 1) and (1, 0). (b) After subtracting the mean, the two perturbations are (-1/2, 1/2) and (1/2, -1/2).

innovation (“observation” minus 12-h control forecast) sample variance for wind and temperature at each observation site by averaging all the squared 12-h innovations in the summer of 2000 at each observation site. Then we choose the smallest wind and temperature innovation sample variance of all observation sites as the observation error variance [please refer to section 3 in Wang and Bishop (2003) for more details].

As shown above, each initial ETKF perturbed member is equally likely. Thus the weights assigned to each of the K perturbed members when calculating ensemble mean and ensemble covariance are the same [$1/K$ in Eqs. (2) and (3)]. However, the probability density of the initial analysis is generally different from that of the perturbed members. Thus, the control member should be weighted differently than the perturbed ensemble members when estimating the mean and covariance of the distribution. The weight that should be assigned to the control member depends on the knowledge of the forecast error covariances of the control forecasts and the individual perturbed ensemble members. This topic needs extra exploration and will be included in future work. Since in this paper we focus on exploring the skill of two ensemble centering schemes, for simplicity we assign zero weight to the control member when calculating ensemble mean and covariance, which will not affect the qualitative comparison results.

4. Comparison of spherical simplex ETKF with symmetric positive–negative paired ETKF

In this section, we compare the performance of the spherical simplex ETKF and the symmetric positive–negative paired ETKF ensembles. Because the experiment results of the spherical simplex ETKF corresponding to solutions 1 and 2 in appendix C are similar, we only show those corresponding to solution 1. Note that solution 1 is a trivial extension of the one-sided ETKF in Eq. (5); hence, it is easy for those familiar with the ETKF to understand.

a. Maintenance of variance along orthogonal basis vectors

As discussed in appendix A, the short-term error covariance estimates by 16 ensemble members in predicting the true mean and true error covariance have rank 15 for the spherical simplex ETKF scheme, but only 8 for the symmetric positive–negative paired ETKF scheme [see second-order terms in (A18) and (A19)]. This expectation is confirmed by the seasonally averaged eigenvalue spectra for 12-h ensemble-based error covariance matrix in observation space in Fig. 2 [see similar plot and definition of the eigenvalue spectra in Fig. 5 of Wang and Bishop (2003)]. Note as discussed in section 3, in our experiment 16 perturbed 12-h forecasts are used to calculate the 12-h ensemble covariance.⁴ Figure 2 shows that while the 12-h ensemble forecast variance for the spherical simplex ETKF ensemble is evenly spread in 15 directions, almost all en-

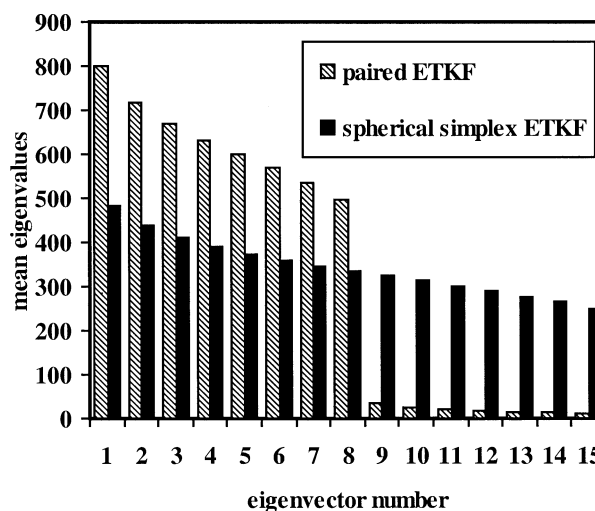


FIG. 2. Seasonally averaged spectra of eigenvalues of 12-h ensemble-estimated forecast error covariance matrices normalized by observation error covariance in observation sites for spherical simplex ETKF and paired ETKF ensembles. The 12-h ensemble covariance is calculated by the 16 perturbed 12-h forecasts.

semble variance is maintained in only 8 directions for the symmetric positive–negative paired ETKF ensemble. As a consequence, short-term optimal growth (Wang and Bishop 2003 section 6; Farrell 1988, 1989) within the ensemble perturbation subspace is larger for the spherical simplex ETKF than for the paired ETKF (not shown).

b. Comparison of initial ensemble variance

Figure 3 shows the square root of the seasonally and vertically averaged initial wind error variance estimated by the spherical simplex ETKF and the symmetric positive–negative paired ETKF ensembles. For both the spherical simplex ETKF and the paired ETKF ensembles, the initial ensemble variance over the ocean is generally larger than that over the land at the same latitude, which is consistent with the fact that rawinsonde observations are more numerous over the land. The spherical simplex ETKF initial ensemble variance over the Southern Hemisphere (SH) is much larger than over the Northern Hemisphere (NH), which is consistent with the fact that the rawinsonde distribution is much less distributed in the SH than in the NH. Another possible reason for this SH–NH contrast is that the spherical simplex ETKF ensemble subspace may properly extract the maximal growing modes of the winter hemisphere (SH in the current experiment) tropospheric wind field. Note the growth of these modes is larger in the winter hemisphere than in the summer hemisphere (NH in the current experiment). In comparison, this NH–SH contrast in the initial ensemble variance is smaller for the paired ETKF than for the spherical simplex ETKF. These results may indicate that the spherical simplex ETKF ensemble represents geographical variations in

⁴ Note that for an ensemble with initial perturbations centering on the analysis, short-term (e.g., 12 h for T42 CCM3) ensemble mean is close to the short-term control forecast. Ensemble subspace rank of the short-term ensemble perturbations with the control run included is approximately the same as that with no control run included.

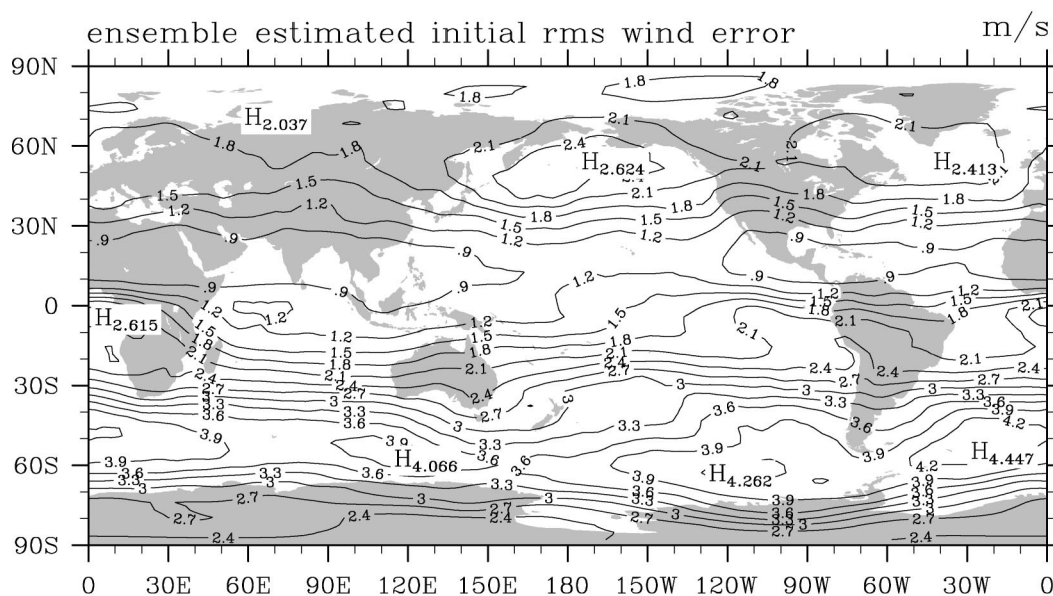
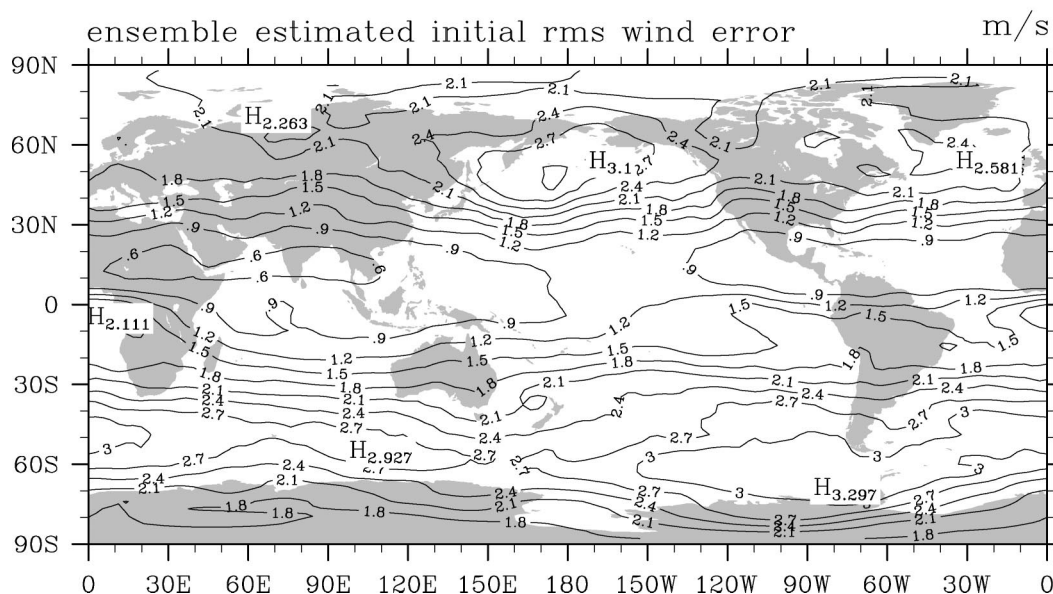
(a) spherical simplex ETKF**(b) paired ETKF**

FIG. 3. Square root of seasonally and vertically averaged ensemble wind variance of initial ensemble perturbations for (a) spherical simplex ETKF ensemble and (b) paired ETKF ensemble. Contour interval is 0.3 m s^{-1} . Label H indicates local maximum.

the analysis error variance due to geographical variations in both observation distribution and error growth rates better than the paired ETKF ensemble. Results (not shown) from runs for NH winter are also consistent with this hypothesis. It is shown from the NH winter runs that the spherical simplex ETKF initial ensemble variance over the NH ocean is larger than that over the SH ocean, and its initial ensemble variance over the NH

ocean is larger than that over the NH continent. In comparison, these contrasts are smaller in the paired ETKF initial ensemble variance.

To better reveal how ensemble spread is governed by the observation density, we plot the rescaling factor that is defined as the ratio of ensemble-estimated initial rms wind error over ensemble-estimated 12-h forecast rms wind error. Such maps give a representation of the geo-

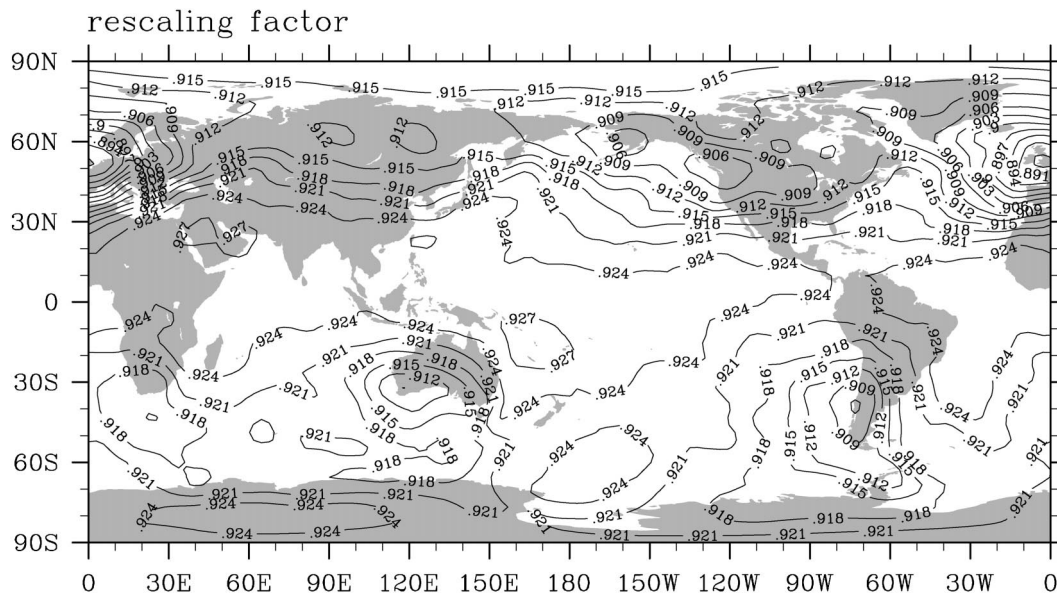
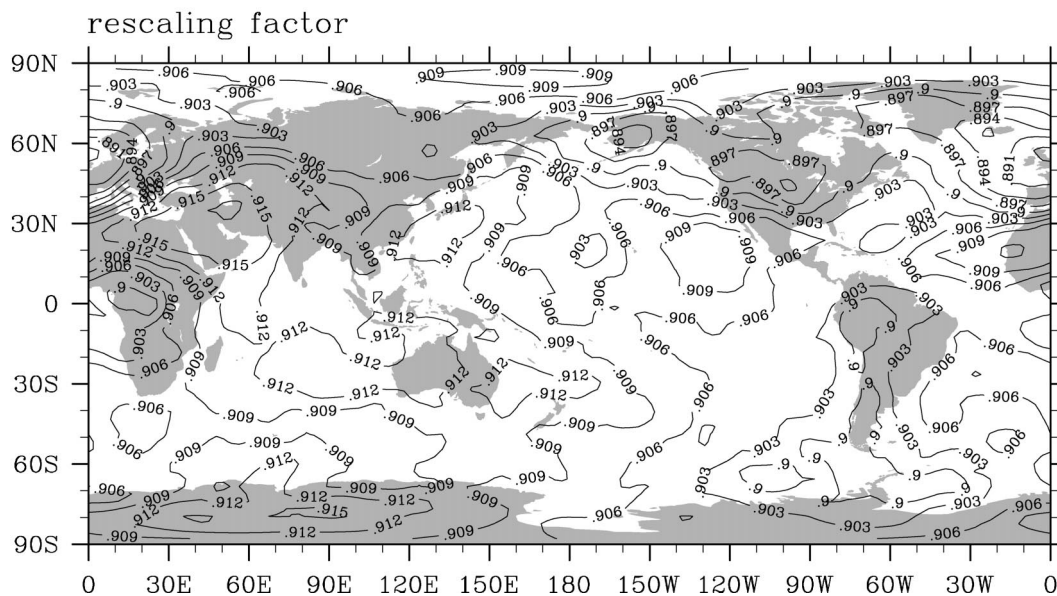
(a) spherical simplex ETKF**(b) paired ETKF**

FIG. 4. Seasonally and vertically averaged ratio of ensemble estimated rms initial wind error over ensemble estimated rms 12-h forecast wind error for (a) spherical simplex ETKF and (b) paired ETKF ensembles. Contour interval is 0.003.

graphical distribution of the factor that rescales 12-h forecast ensemble spread into initial ensemble spread. In regions where forecast error variance is large but observations are populated, the rescaling factor is expected to be small [see also the discussion in section 4 of Wang and Bishop (2003)]. Figure 4 shows the vertically and seasonally averaged rescaling factor. The effective rescaling factor for the spherical simplex ETKF not only

reflects the high concentration of observations over Europe and North America, it is also able to account for the smaller midlatitude observation concentrations over Southern Hemisphere (SH) continents. In contrast, the rescaling factor of the positive-negative paired ETKF does not account for these land-based observation concentrations within SH as well as the spherical simplex ETKF.

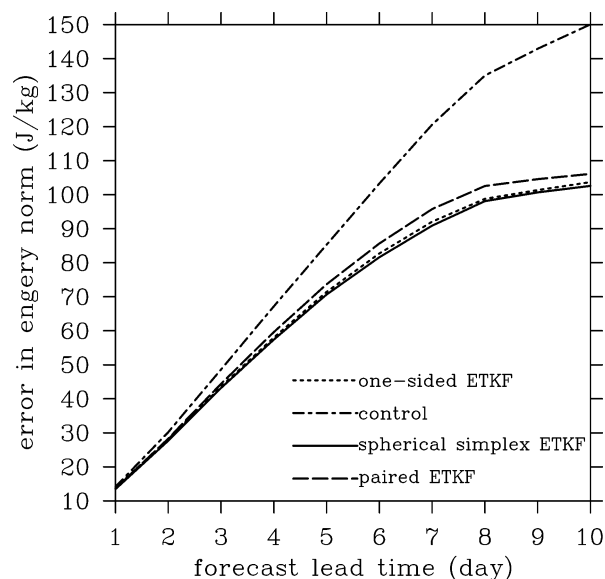


FIG. 5. Globally (all model grids at 200, 500, and 850 hPa) and seasonally averaged ensemble mean forecast error in terms of the approximate total energy norm as a function of forecast lead time for the spherical simplex ETKF, paired ETKF, and one-sided ETKF ensembles. The corresponding measurement of the control forecast error is also shown.

c. Root-mean-square error of the ensemble mean

Figure 5 shows 200-, 500- and 850-hPa globally averaged ensemble mean forecast error in terms of the approximate energy norm [see definition in Eq. (26) of Wang and Bishop (2003)] for the spherical simplex ETKF, paired ETKF, and one-sided ETKF ensembles with 16 perturbed members each (recall that as discussed in section 3 the control is not included when calculating the ensemble mean for any of the ensembles considered here including the one-sided ETKF). The corresponding measurements of control forecast errors are also shown for comparison. The verifications are the NCEP–NCAR reanalysis data. Figure 5 shows that the ensemble mean of the spherical simplex ETKF is more accurate than the symmetric positive–negative paired ETKF throughout 1–10-day forecast lead times. Further analysis demonstrates that the improvement of the ensemble mean from spherical simplex ETKF relative to that from the paired ETKF is mainly in the SH, the winter hemisphere. Extra runs for the time of NH winter show statistically significant improvement of the spherical simplex ETKF relative to the paired ETKF on both hemispheres. From Fig. 5, although the paired ETKF is centered on the analysis initially, its ensemble mean is less accurate than that of the one-sided ETKF from 2–10-day forecast lead times. Note that for the same ensemble size the one-sided ETKF ensemble has one more subspace rank than the spherical simplex ETKF. All these results indicate that for a given ensemble size it is more important to represent as many error directions as possible than to maintain a little more accuracy in fewer directions.

Figure 5 also shows that there is a small improvement of the spherical simplex ETKF ensemble mean over the one-sided ETKF ensemble mean at all lead times. We speculate three reasons that may explain why in our current experiment the ensemble mean of the one-sided ETKF is only a little less accurate than that of the spherical simplex ETKF. First, the one-sided ETKF has one more subensemble direction than the spherical simplex ETKF for a given ensemble size. Second, the initial ensemble variance is distributed evenly onto 16 orthogonal directions spanned by the one-sided ETKF initial perturbations. Thus the deviation of the initial one-sided ETKF ensemble mean from the analysis is smaller than that in the case in which the same amount of initial ensemble variance is contained mainly in one direction. Third, because of suboptimality in the data assimilation scheme used to produce the reanalysis datasets, the analysis may be a relatively poor approximation to an optimal minimum error variance estimate. Such suboptimality decreases the advantages of centering the initial perturbations on the analysis rather than on some point close to the analysis (note from the second explanation above that the one-sided ETKF initial ensemble is centered on a point close to the analysis).

d. Comparison of ensemble predictions of innovation variance

To compare the skill of the ensemble spread in predicting the forecast error variance, we adopt the methods introduced in section 8 of Wang and Bishop (2003). The results below show that for 1- and 2-day forecast lead times the skill of the ensemble predictions of the innovation variance from the spherical simplex ETKF significantly outperforms that of the paired ETKF. For longer forecast lead times from 3 to 10 days, their skills become close.

Figure 6 shows the relationship between the sample innovation variance and the ensemble variance for 500-hPa U at 1-day forecast lead time. This figure is generated by first drawing a scatterplot for which the ordinate and abscissa of each point is respectively given by the squared 500-hPa U wind innovation and 500-hPa U wind ensemble variance at 1-day forecast at one midlatitude rawinsonde observation location. The innovation is defined here as the difference between the verifying analysis and the 1-day ensemble mean forecast at the rawinsonde observation sites. Points collected correspond to all midlatitude stations and all 1-day 500-hPa U forecasts throughout the NH summer in year 2000. To begin, we divide the points into four equally populated bins, arranged in order of increasing ensemble variance. Then we average the squared innovation and ensemble variance in each bin, respectively. Connecting the averaged points then yields a curve describing the relationship between the sample innovation variance and the ensemble variance. The results corresponding to the 4-bin and 32-bin cases for 1-day forecast lead time are

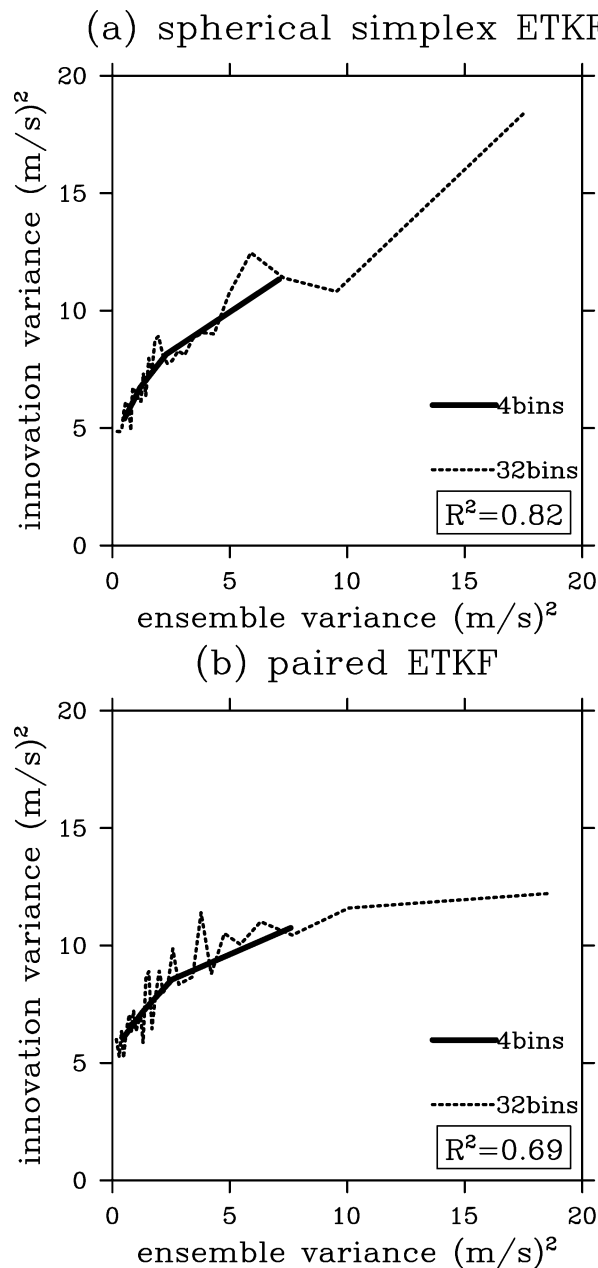


FIG. 6. The relationship between the 500-mb U sample innovation variance and the ensemble variance at 1-day forecast lead time. Solid lines are for the 4-bin case and dashed lines for 32-bin case. The R^2 value is a measurement of how noisy the dashed curve is relative to the solid curve. For details please refer to the text in section 4d.

shown in Fig. 6. First, note that the range of innovation variance resolved by the spherical simplex ETKF ensemble variance is much larger than that of the paired ETKF. A statistical test based on halving the data size was used to confirm this result. Second, as the sample size in each bin is decreased (e.g., from 4-bin case to 32-bin case), the relationship between sample innovation variance and the ensemble variance for 1-day forecasts becomes noisier for the paired ETKF than for the

spherical simplex ETKF. The noisiness of the dashed curve relative to the solid curve is measured by the R^2 value (Ott 1993). Less noisiness corresponds to large R^2 value. A statistical t test (Ott 1993) is used to confirm that the R^2 value of the spherical simplex ETKF is significantly larger than that of the paired ETKF. According to the analysis in section 8 of Wang and Bishop (2003), these results show that for 1-day forecast (true for 2-day forecast as well; not shown), the ensemble spread of the spherical simplex ETKF is more accurate in predicting the forecast error variance than that of the paired ETKF. Our results also shows that for longer forecast lead times from 3 to 10 days, the skills of the two centering schemes in predicting the innovation variance become statistically indistinguishable. Figure 7, which shows the results for 10-day forecast lead time, illustrates this point.

5. Summary and discussion

In this paper, we tested the performance of two ensemble-centering methods for the ETKF ensemble. One was the common symmetric positive-negative paired centering and the other was the spherical simplex centering. In the spherical simplex scheme, one more perturbation was added to one-sided ETKF initial perturbations such that (a) the sum of the new set of initial perturbations equaled zero, (b) the sample covariance of the new perturbations was equal to the ETKF estimated analysis error covariance matrix, and (c) all the new initial perturbations were equally likely.

For an ensemble of K perturbed members, the spherical simplex ETKF maintained comparable amounts of variance in $K - 1$ orthogonal and uncorrelated directions as compared to only $K/2$ directions for the paired ETKF over short forecast lead times. The initial ensemble variance from the spherical simplex ETKF better reflected the geographical variations of the observations than the paired ETKF. The spherical simplex ETKF ensemble mean was found to be more accurate than the mean of the positive-negative paired ETKF ensemble. The spherical simplex ETKF ensemble variance resolved a significantly larger range of sample innovation variance than the paired ETKF for 1- and 2-day forecast lead times. Because the spherical simplex ETKF initial perturbations were generated by simply postmultiplying the one-sided ETKF initial perturbations by a $(K - 1) \times K$ matrix (section 2c), where K equals the number of perturbed members, the computational expense of generating the spherical simplex ETKF ensemble is about the same as that of generating the symmetric positive-negative paired ETKF ensemble.

In section 2, it was algebraically demonstrated that the subtract-mean centering scheme proposed by Toth and Kalnay (1997) does not preserve the ETKF estimated analysis error covariance. For this reason, we did not apply it to the ETKF ensemble. However, intrigued by Toth and Kalnay's (1997) findings, we went ahead

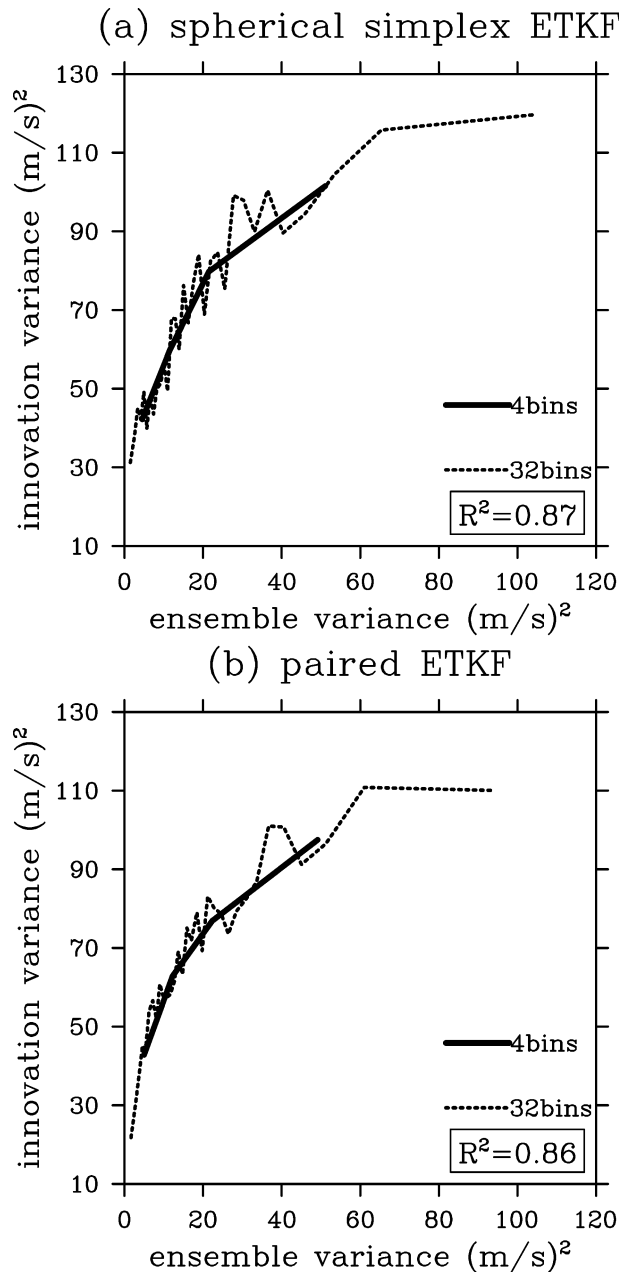


FIG. 7. As in Fig. 6, except for 10-day forecast lead time.

and tested the subtract-mean centering for the breeding ensembles [see section 5b of Toth and Kalnay (1997)]. The experimental environment was the same as for the ETKF (see also Wang and Bishop 2003). We found that the symmetric positive-negative paired breeding had inferior forecast skill in both mean and ensemble spread than the subtract-mean breeding. Examination of the eigenvalue spectra of the 12-h ensemble forecast covariance showed that the $K/2$ (8 in this experiment) trailing eigenvalues of the subtract-mean breeding ensemble were significantly larger than those of the paired

breeding. We thus speculated that it was the ability of the subtract-mean centering to maintain error variance in more than $K/2$ directions that made the subtract-mean breeding perform better than the paired breeding. However, these results appear to be inconsistent with those reported in Toth and Kalnay (1997) in which it was found that the paired breeding was more skillful than the subtract-mean breeding. We have no firm explanation for this discrepancy.

When initial perturbations are generated to have equal probability density they should be assigned equal weights when calculating ensemble means and covariances. However, in most circumstances, the control forecast and/or the members of a multimodel ensemble are unlikely to be equally probable. In such circumstances, one should assign different weights to ensemble members when computing means and covariances. We shall address this issue in more detail in future work.

Acknowledgments. The authors gratefully acknowledge financial support from ONR Grant N00014-00-1-0106 and ONR Project Element 0601153N with Project Number BE-033-0345. We also greatly appreciate the insightful comments from the two reviewers.

APPENDIX A

Theoretical Analysis on the Skill of Sigma Points in Predicting Mean and Covariance

Assuming the initial state vector is a symmetrically distributed random vector \mathbf{x} with mean $\bar{\mathbf{x}}$ and covariance \mathbf{P}_x , an important aim of ensemble forecasting is to predict the mean $\bar{\mathbf{y}}$ and covariance \mathbf{P}_y of the distribution of the forecast states given a prediction model and observations. Let \mathbf{y} denote a random draw from this distribution, where \mathbf{y} is related to \mathbf{x} by the nonlinear model operator M :

$$\mathbf{y} = M(\mathbf{x}). \quad (\text{A1})$$

The basic idea of the sigma-point ensemble is to find initial perturbations that capture the mean and covariance information while at the same time permitting the direct propagation of the information through the nonlinear model operator. Specifically, K initial state vectors \mathbf{x}_i , $i = 1, \dots, K$, are chosen to reflect the first two moments of \mathbf{x} , that is,

$$\sum_{i=1}^K w_i \mathbf{x}_i = \bar{\mathbf{x}}, \quad (\text{A2})$$

$$\sum_{i=1}^K w_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = \mathbf{P}_x, \quad (\text{A3})$$

where w_i , $i = 1, \dots, K$, are weights satisfying $\sum_{i=1}^K w_i = 1$. If \mathbf{x}_i , $i = 1, \dots, K$, are equally likely, then w_i , $i = 1, \dots, K$, are equal to each other. Equation (A3) is satisfied if $\sqrt{w_i}(\mathbf{x}_i - \bar{\mathbf{x}})$ is chosen to be equal to the i th column of the square root matrix of \mathbf{P}_x . The initial per-

turbations constructed in this way are called sigma points, which are denoted as $\sigma_i = \mathbf{x}_i - \bar{\mathbf{x}}$. Then Eqs. (A2) and (A3) become

$$\sum_{i=1}^K w_i \sigma_i = \mathbf{0}, \quad (\text{A4})$$

$$\sum_{i=1}^K w_i \sigma_i (\sigma_i)^T = \mathbf{P}_x. \quad (\text{A5})$$

To estimate the mean and covariance of the future state, each point \mathbf{x}_i is propagated through (A1) to obtain $\mathbf{y}_i = M(\mathbf{x}_i)$. Then, the estimated mean and covariance of the forecast state \mathbf{y} are given by

$$\bar{\mathbf{y}} = \sum_{i=1}^K w_i \mathbf{y}_i, \quad (\text{A6})$$

$$\mathbf{P}_y = \sum_{i=1}^K w_i (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T. \quad (\text{A7})$$

The ETKF ensemble generation scheme (Bishop et al. 2001; Wang and Bishop 2003) together with other forms of the deterministic ensemble square root filter (Tippett et al. 2003; Anderson 2001; Whitaker and Hamill 2002) produce sigma-point ensembles aiming to satisfy (A4) and (A5). Arguably, the singular vector ensemble forecast scheme (Molteni et al. 1996) at ECMWF and the breeding scheme (Toth and Kalnay 1993, 1997) at NCEP can also be regarded as approximations to sigma-point ensembles. In the following, we perform a multidimensional Taylor series in order to analyze the skill of sigma points in predicting the true mean and true covariance (see also Julier and Uhlmann 2002b). The discussion is first based on the assumption that \mathbf{P}_x is precisely modeled by (A2)–(A5). The implications of the fact that the ensemble size is too small to model \mathbf{P}_x correctly is discussed at the end of section A2.

a. True mean and covariance

Consider the initial random vector \mathbf{x} as the mean $\bar{\mathbf{x}}$ plus a zero-mean disturbance $\Delta\mathbf{x}$ with covariance \mathbf{P}_x . Then the Taylor series expansion of the nonlinear transformation $M(\mathbf{x})$ about $\bar{\mathbf{x}}$ is

$$\begin{aligned} \mathbf{y} &= M(\bar{\mathbf{x}} + \Delta\mathbf{x}) \\ &= M(\bar{\mathbf{x}}) + D_{\Delta\mathbf{x}}M + \frac{D_{\Delta\mathbf{x}}^2M}{2!} + \frac{D_{\Delta\mathbf{x}}^3M}{3!} + \frac{D_{\Delta\mathbf{x}}^4M}{4!} + \dots, \end{aligned} \quad (\text{A8})$$

where the $D_{\Delta\mathbf{x}}M$ operator evaluates the total differential of M when perturbed around $\bar{\mathbf{x}}$ by $\Delta\mathbf{x}$. The operator can be written as

$$D_{\Delta\mathbf{x}} = \sum_{k=1}^n \Delta x_k \frac{\partial}{\partial x_k} = (\Delta\mathbf{x})^T \cdot \nabla, \quad (\text{A9})$$

which acts on M on a component-by-component basis. Here, n is the number of elements in \mathbf{x} . The j th term in (A8) is given by

$$\begin{aligned} \frac{D_{\Delta\mathbf{x}}^j M}{j!} &= \frac{D_{\Delta\mathbf{x}} \{ D_{\Delta\mathbf{x}} [\dots (D_{\Delta\mathbf{x}} M)] \}}{j!} \\ &= \frac{1}{j!} \left(\sum_{k=1}^n \Delta x_k \frac{\partial}{\partial x_k} \right)^j M(\mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}}. \end{aligned} \quad (\text{A10})$$

Equation (A10) can be expressed as a sum of components, each of which is a product of a j th-order product of $\Delta\mathbf{x}$ and a j th-order differential of M with respect to \mathbf{x} .

Recall that $\Delta\mathbf{x}$ is assumed to be symmetrically distributed with its mean equal to zero. By symmetry, the expected value of all odd terms in (A8) is zero. Taking expectations (denoted by operator E) of (A8), we obtain the true mean (denoted by the superscript t)

$$\bar{\mathbf{y}}^t = M(\bar{\mathbf{x}}) + \left(\frac{\nabla^T \mathbf{P}_x \nabla}{2!} \right) M + E \left(\frac{D_{\Delta\mathbf{x}}^4 M}{4!} + \dots \right). \quad (\text{A11})$$

The second-order even terms in (A11) is obtained by noting that

$$\begin{aligned} E \left(\frac{D_{\Delta\mathbf{x}}^2 M}{2!} \right) &= E \left[\left(\frac{\Delta\mathbf{x}^T \nabla \Delta\mathbf{x}^T \nabla}{2!} \right) M \right] \\ &= E \left[\left(\frac{\nabla^T \Delta\mathbf{x} \Delta\mathbf{x}^T \nabla}{2!} \right) M \right] = \left(\frac{\nabla^T \mathbf{P}_x \nabla}{2!} \right) M. \end{aligned} \quad (\text{A12})$$

The true covariance of the forecast state is given by

$$\mathbf{P}_y^t = E[(\mathbf{y} - \bar{\mathbf{y}}^t)(\mathbf{y} - \bar{\mathbf{y}}^t)^T]. \quad (\text{A13})$$

Substitute (A8) and (A11) into (A13). Recall the symmetry of $\Delta\mathbf{x}$ and thus the expected value of all odd order terms of $\Delta\mathbf{x}$ evaluates to zero. The true covariance (written up to fourth order) is

$$\mathbf{P}_y^t = \mathbf{M} \mathbf{P}_x \mathbf{M}^T + E \left[\frac{D_{\Delta\mathbf{x}} M (D_{\Delta\mathbf{x}}^3 M)^T}{3!} + \frac{D_{\Delta\mathbf{x}}^2 M (D_{\Delta\mathbf{x}}^2 M)^T}{2! \times 2!} + \frac{D_{\Delta\mathbf{x}}^3 M (D_{\Delta\mathbf{x}} M)^T}{3!} \right] - \left[\left(\frac{\nabla^T \mathbf{P}_x \nabla}{2!} \right) M \right] \left[\left(\frac{\nabla^T \mathbf{P}_x \nabla}{2!} \right) M \right]^T + \dots, \quad (\text{A14})$$

where \mathbf{M} is the Jacobian matrix of M or the linearized dynamics operator of M (Ide et al. 1997) evaluated about

$\bar{\mathbf{x}}$, that is, $D_{\Delta\mathbf{x}} M = \mathbf{M} \Delta\mathbf{x}$. Note the model is assumed to be perfect in the above discussion.

b. Sigma-point ensemble mean and covariance

The sigma-point ensemble predicts the mean from the propagated set of points \mathbf{y}_i by using (A6). Consider the Taylor series for the transition of each point \mathbf{x}_i . Each propagated point \mathbf{y}_i can be expressed as the Taylor series about $\bar{\mathbf{x}}$,

$$\begin{aligned} \mathbf{y}_i &= M(\mathbf{x}_i) = M(\bar{\mathbf{x}} + \boldsymbol{\sigma}_i) \\ &= M(\bar{\mathbf{x}}) + D_{\sigma_i}M + \frac{D_{\sigma_i}^2 M}{2!} + \frac{D_{\sigma_i}^3 M}{3!} + \frac{D_{\sigma_i}^4 M}{4!} + \dots \end{aligned} \quad (\text{A15})$$

Applying Eq. (A6), the predicted mean is

$$\begin{aligned} \bar{\mathbf{y}} &= \sum_{i=1}^K w_i \left[M(\bar{\mathbf{x}}) + D_{\sigma_i}M + \frac{D_{\sigma_i}^2 M}{2!} + \frac{D_{\sigma_i}^3 M}{3!} \right. \\ &\quad \left. + \frac{D_{\sigma_i}^4 M}{4!} + \dots \right]. \end{aligned} \quad (\text{A16})$$

Recall from (A9) and (A12) that

$$D_{\sigma_i}M = \boldsymbol{\sigma}_i^T \cdot \nabla M, \quad \text{and}$$

$$\frac{D_{\sigma_i}^2 M}{2!} = \left(\frac{\nabla^T \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i^T \nabla}{2!} \right) M, \quad (\text{A17})$$

and apply Eqs. (A4) and (A5) on the second and third terms of Eq. (A16), Eq. (A16) becomes

$$\begin{aligned} \bar{\mathbf{y}} &= M(\bar{\mathbf{x}}) + \left(\frac{\nabla^T \mathbf{P}_x \nabla}{2!} \right) M \\ &\quad + \sum_{i=1}^K w_i \left(\frac{D_{\sigma_i}^3 M}{3!} + \frac{D_{\sigma_i}^4 M}{4!} + \dots \right). \end{aligned} \quad (\text{A18})$$

To obtain the sigma-point counterpart of (A14) for the second-moment of the distribution, substitute (A15) and (A18) into (A7) and recall (A5) and the identity $D_{\sigma_i}M = \mathbf{M}\boldsymbol{\sigma}_i$, the sigma-point predicted covariance (written up to fourth order) is

$$\begin{aligned} \mathbf{P}_y &= \mathbf{M}\mathbf{P}_x\mathbf{M}^T + \sum_{i=1}^K w_i \left[D_{\sigma_i}M \left(\frac{D_{\sigma_i}^2 M}{2!} \right)^T + \frac{D_{\sigma_i}^2 M}{2!} (D_{\sigma_i}M)^T \right] + \sum_{i=1}^K w_i \left[\frac{D_{\sigma_i}M (D_{\sigma_i}^3 M)^T}{3!} + \frac{D_{\sigma_i}^2 M (D_{\sigma_i}^2 M)^T}{2! \times 2!} + \frac{D_{\sigma_i}^3 M (D_{\sigma_i}M)^T}{3!} \right] \\ &\quad - \left[\left(\frac{\nabla^T \mathbf{P}_x \nabla}{2!} \right) M \right] \left[\left(\frac{\nabla^T \mathbf{P}_x \nabla}{2!} \right) M \right]^T - \sum_{i=1}^K w_i \left[D_{\sigma_i}M \left(\sum_{i=1}^K w_i \frac{D_{\sigma_i}^3 M}{3!} \right)^T \right. \\ &\quad \left. + \left(\sum_{i=1}^K w_i \frac{D_{\sigma_i}^3 M}{3!} \right) (D_{\sigma_i}M)^T \right] + \dots \end{aligned} \quad (\text{A19})$$

Comparing (A19) with (A14) and (A18) with (A11), for a sufficiently large sigma-point ensemble that can represent all uncertain directions of the system so that the analysis error covariance can be modeled precisely by (A3), the sigma-point ensemble mean and covariance agree with the true mean and true covariance at least up to the second-order term. For symmetric positive-negative paired sigma points with equal weights on each member of the same pair, by symmetry the odd order terms in (A18) and (A19) vanish, and therefore the accuracy of the estimated mean and covariance is up to third order. For the simplex sigma-point ensemble, the $\boldsymbol{\sigma}_i$'s are not symmetric, and hence the ensemble mean and covariance only have second-order accuracy.

The above analysis is predicated upon the assumption that the analysis error covariances were perfectly modeled by (A3). However, in reality there is no prior knowledge about the true analysis error covariance, and the ensemble size K is much smaller than the rank of the true analysis error covariance. In this situation, the sigma-point-estimated \mathbf{P}_x , and thus the second-order term on the right-hand side of (A18) and (A19), has rank of $K - 1$ for the simplex ensemble, but only $K/2$ for the symmetric posi-

tive-negative paired ensemble. In other words, the simplex sigma-point ensemble can describe error variance in more directions than the paired sigma-point ensemble. In the case where the initial sigma points precisely model the analysis error variance for each direction within the ensemble subspace, the simplex sigma point would provide second-order-accurate mean and covariance in $K - 1$ directions whereas the paired sigma point would have third-order accuracy in only $K/2$ directions.

Note also if the sum of the initial perturbations is not zero and \mathbf{P}_x is estimated by the outer product of uncentered perturbations such as those from the one-sided ETKF, the ensemble mean and ensemble covariance calculated by (A6) and (A7) will not even agree with the first- and second-order terms of the Taylor expansions for the true mean and covariance.

APPENDIX B

Sense in which One-Sided ETKF Initial Perturbations are Equally Likely

To see the sense in which one-sided ETKF initial perturbations are equally likely, note that for an n -di-

mensional, normally distributed, and zero-mean random variable \mathbf{x}' , the pdf, that is, $p(\mathbf{x}')$, is given by,

$$p(\mathbf{x}') = (2\pi)^{-n/2} (\det \mathbf{P})^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{x}')^T \mathbf{P}^{-1} (\mathbf{x}') \right], \quad (\text{B1})$$

where \mathbf{P} is the covariance matrix of \mathbf{x}' and $\det \mathbf{P}$ is the determinant of \mathbf{P} (see Dee 1995; Ross 1998). In the following we show that assuming a multidimensional normal distribution, the one-sided ETKF initial perturbations projected onto observation space and normalized by the root-mean-square observation error have equal pdfs. See denotations in section 2a.

Define $\mathbf{E} = \tilde{\mathbf{H}} \mathbf{Z}^f \mathbf{C} \mathbf{F}^{-1/2}$, where $\tilde{\mathbf{H}}$ is the observation operator \mathbf{H} normalized by the square root of the observation error covariance matrix \mathbf{R} , that is, $\tilde{\mathbf{H}} = \mathbf{R}^{-1/2} \mathbf{H}$. Then keeping in mind that (1) postmultiplied by the last column of \mathbf{C} is a zero vector, it is easy to verify that

$$\tilde{\mathbf{H}} \mathbf{P}^a \tilde{\mathbf{H}}^T = \mathbf{E} \mathbf{E}^T. \quad (\text{B2})$$

Here (B2) can be regarded as an approximate eigenstructure expression of $\tilde{\mathbf{H}} \mathbf{P}^a \tilde{\mathbf{H}}^T$ since $\mathbf{E}^T \mathbf{E} = \mathbf{I}$ but $\mathbf{E} \mathbf{E}^T \neq \mathbf{I}$. Defining $\mathbf{D} = \mathbf{F}(\mathbf{F} + \mathbf{I})^{-1}$, then from (5), (6), and the definition of \mathbf{E} and \mathbf{Z}^f , we obtain

$$\tilde{\mathbf{H}} \mathbf{X}^a = \tilde{\mathbf{H}}(\mathbf{x}_1^a, \mathbf{x}_2^a, \dots, \mathbf{x}_{K-1}^a) = \sqrt{K} \mathbf{E} \mathbf{D}^{1/2}, \quad (\text{B3})$$

$$\tilde{\mathbf{H}} \mathbf{P}^a \tilde{\mathbf{H}}^T = \mathbf{E} \mathbf{D} \mathbf{E}^T. \quad (\text{B4})$$

Similarly the matrices \mathbf{E} and \mathbf{D} are the approximate eigenvector and eigenvalue matrices of $\tilde{\mathbf{H}} \mathbf{P}^a \tilde{\mathbf{H}}^T$. Replace \mathbf{P}^{-1} in (B1) by the approximate inversion^{B1} of $\tilde{\mathbf{H}} \mathbf{P}^a \tilde{\mathbf{H}}^T$, which according to (B4) is $(\tilde{\mathbf{H}} \mathbf{P}^a \tilde{\mathbf{H}}^T)^{-1} \approx \mathbf{E} \mathbf{D}^{-1} \mathbf{E}^T$. Also replace \mathbf{x}' in (B1) by $\tilde{\mathbf{H}} \mathbf{x}_i^a$, i from 1 to $K-1$ in (B3). Then it is easy to verify that the probability density of $\tilde{\mathbf{H}} \mathbf{x}_i^a$, that is, $p(\tilde{\mathbf{H}} \mathbf{x}_i^a)$, is the same for all i . Hence, these perturbations are equally likely.

APPENDIX C

Spherical Simplex Sigma Points

The minimum number of sigma points required to have the same mean and covariance as an n -dimensional random variable is $n+1$ (Julier and Uhlmann 2002a; Julier 2003). Specifically, to satisfy (A4) and (A5), the minimum number of σ_i is $n+1$ if \mathbf{x} is an n -dimensional random variable. We call any set of such points *simplex sigma points*. Besides satisfying (A4) and (A5), the *spherical* simplex sigma points also require that the $n+1$ σ_i 's are equally likely. If the covariance matrix of \mathbf{x} , \mathbf{P}_x , is homogeneous, that is, the eigenvalues of \mathbf{P}_x are all equal to each other, then it is easy to verify from Eq. (B1) that in order for each σ_i to be equally likely (i.e., have the same probability density), each σ_i must have equal distance to the origin. In other words, the $n+1$ σ_i 's lie on a *hypersphere*.

Recall in section 2c, for $K-1$ one-sided independent ETKF initial perturbations, we seek a $(K-1) \times K$

matrix \mathbf{U} to satisfy (a) $\mathbf{U} \mathbf{1} = \mathbf{0}$, (b) $\mathbf{U} \mathbf{U}^T = \mathbf{I}$, and (c) each column of \mathbf{U} has the same magnitude. In condition (a), $\mathbf{0}$ is a vector with each element equal to zero and $\mathbf{1}$ is a vector with each element equal to one. From these three requirements and the definition of the spherical simplex points above, solving for the $(K-1) \times K$ matrix \mathbf{U} is equivalent to obtaining K spherical simplex sigma points, each of which is a $(K-1)$ -dimensional variable. The mean of these K spherical simplex sigma points is a zero vector, the covariance is proportional to the identity matrix, and the distance of these K points to the origin (or in other words the magnitude of each point) must be equal to each other. In our experiment, we construct two easy sets of spherical simplex points \mathbf{U} that satisfy conditions (a)–(c).

a. Solution 1

Here, we show how a trivial extension of the ETKF can be used to obtain ensemble satisfying conditions (a)–(c). We will show below that the $(K-1) \times K$ matrix \mathbf{E}^T , the transpose of \mathbf{E} that is defined in section 2a, satisfies the three conditions. Then from (5) and (7), the spherical simplex ETKF initial perturbations are

$$\mathbf{Y}^a = \mathbf{X}^f \mathbf{E} (\mathbf{F} + \mathbf{I})^{-1/2} \mathbf{E}^T. \quad (\text{C1})$$

First we prove \mathbf{E}^T satisfies condition (a), that is, $\mathbf{E}^T \cdot \mathbf{1} = \mathbf{0}$. Assume we have K forecast perturbations defined as in (1). From (1),

$$\mathbf{Z}^f \mathbf{1} = \mathbf{0}, \quad (\text{C2})$$

where $\mathbf{Z}^f = \mathbf{X}^f / \sqrt{K}$. Now, premultiply (C2) by $(\mathbf{H} \mathbf{Z}^f)^T \mathbf{R}^{-1} \mathbf{H}$, where as in Bishop et al. (2001) and Wang and Bishop (2003), \mathbf{H} is the observation operator and \mathbf{R} is the observation error covariance matrix, then

$$(\mathbf{H} \mathbf{Z}^f)^T \mathbf{R}^{-1} \mathbf{H} \mathbf{Z}^f \mathbf{1} = \mathbf{0}. \quad (\text{C3})$$

Perform eigenvector decomposition on $(\mathbf{H} \mathbf{Z}^f)^T \mathbf{R}^{-1} \mathbf{H} \mathbf{Z}^f$, then we have

$$\mathbf{C} \mathbf{\Gamma} \mathbf{C}^T \cdot \mathbf{1} = \mathbf{0}, \quad (\text{C4})$$

where as defined in Bishop et al. (2001) and Wang and Bishop (2003) \mathbf{C} and $\mathbf{\Gamma}$ are the eigenvector and eigenvalue matrices of $(\mathbf{H} \mathbf{Z}^f)^T \mathbf{R}^{-1} \mathbf{H} \mathbf{Z}^f$. Both \mathbf{C} and $\mathbf{\Gamma}$ are $K \times K$ matrices. In this discussion and throughout this paper, assume that the eigenvectors and corresponding eigenvalues are organized from left to right in order of decreasing eigenvalues. Note that the lack of linear independence of the columns of \mathbf{Z}^f implied by (C2) mean that the smallest eigenvalue in $\mathbf{\Gamma}$ is equal to zero. Define a new diagonal matrix $\hat{\mathbf{\Gamma}}$ from $\mathbf{\Gamma}$ by setting the last eigenvalue in $\mathbf{\Gamma}$ to one. Then premultiply (C4) by $\hat{\mathbf{\Gamma}}^{-1} \mathbf{C}^T$. We have $\hat{\mathbf{\Gamma}}^{-1} \mathbf{C}^T \cdot \mathbf{C} \mathbf{\Gamma} \mathbf{C}^T \cdot \mathbf{1} = \hat{\mathbf{\Gamma}}^{-1} \mathbf{\Gamma} \mathbf{C}^T \cdot \mathbf{1} = \mathbf{0}$. Thus,

$$\mathbf{E}^T \cdot \mathbf{1} = \mathbf{0}, \quad (\text{C5})$$

where \mathbf{E} , as defined in section 2a, a $K \times (K-1)$ matrix, contains the first $K-1$ columns of \mathbf{C} . So far, we have shown that \mathbf{E}^T satisfies condition (a).

^{B1} Called pseudoinverse in linear algebra (Nakos and Joyner 1998).

Because columns of \mathbf{E} are eigenvectors of $(\mathbf{H}\mathbf{Z}^f)^T \mathbf{R}^{-1} \mathbf{H}\mathbf{Z}^f$, it is easy to verify that

$$\mathbf{E}^T (\mathbf{E}^T)^T = \mathbf{E}^T \mathbf{E} = \mathbf{I}. \quad (\text{C6})$$

So condition (b) is also satisfied by \mathbf{E}^T .

Because the last eigenvalue in $\mathbf{\Gamma}$ is equal to zero and because each column of \mathbf{C} is of unit length, the elements in the last column of \mathbf{C} are equal to $1/\sqrt{K}$. Using this fact and the fact that $\mathbf{C}\mathbf{C}^T = \mathbf{I}$, one may show that

$$[(\mathbf{E}^T)^T \mathbf{E}^T]_{ii} = (\mathbf{E}\mathbf{E}^T)_{ii} = 1 - \frac{1}{K}, \quad i = 1, \dots, K, \quad (\text{C7})$$

that is, the diagonal elements of $\mathbf{E}\mathbf{E}^T$ are all equal to each other and thus each column of \mathbf{E}^T has equal magnitude. So \mathbf{E}^T also satisfies condition (c). Thus (C5), (C6), and (C7) imply that \mathbf{E}^T , which is the first $(K - 1)$ rows of \mathbf{C}^T , satisfies all three requirements.

b. Solution 2^{C1}

Another solution to \mathbf{U} is from purely geometric considerations to satisfy the three conditions. It is obtained by first considering the problem of choosing a set of points that capture the three requirements in a single dimension and then extend the set to two dimensions, three dimensions, and so on. The matrix \mathbf{U} is then in the format as

$$\begin{bmatrix} u_{11} & u_{12} & 0 & 0 & \cdots & 0 \\ u_{21} & u_{22} & u_{23} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{(K-1)1} & u_{(K-1)2} & u_{(K-1)3} & u_{(K-1)4} & \cdots & u_{(K-1)K} \end{bmatrix}, \quad (\text{C8})$$

where u_{ij} is the j th point in the K sigma points for the i th dimension, \mathbf{e}_i . First choose a set of points $u_{11} = (x_1)$ and $u_{12} = (x_2)$ to satisfy the three requirements in a single direction, \mathbf{e}_1 , then x_1 and x_2 need to satisfy

$$x_1 + x_2 = 0, \quad (\text{C9})$$

$$x_1^2 + x_2^2 = 1, \quad (\text{C10})$$

$$x_1^2 = x_2^2. \quad (\text{C11})$$

So, we choose $x_1 = -1/\sqrt{2}$ and $x_2 = 1/\sqrt{2}$. To extend the set to two dimensions, points u_{11} and u_{12} are extended in the direction of \mathbf{e}_2 by $-x_3$, that is, the extended points are $(x_1, -x_3)^T$ and $(x_2, -x_3)^T$. A new point is added at $(0, s_3 x_3)^T$ (see Fig. C1). This extension ensures that the mean and covariance constraints are maintained in the \mathbf{e}_1 direction. Also there is no correlation between \mathbf{e}_1 and \mathbf{e}_2 . The mean and covariance constraints in the \mathbf{e}_2 direction are

$$-x_3 - x_3 + s_3 x_3 = 0, \quad (\text{C12})$$

$$x_3^2 + x_3^2 + (s_3 x_3)^2 = 1. \quad (\text{C13})$$

^{C1} J. Purser at NCEP once independently derived a spherical simplex solution similar to (C15) but never published.

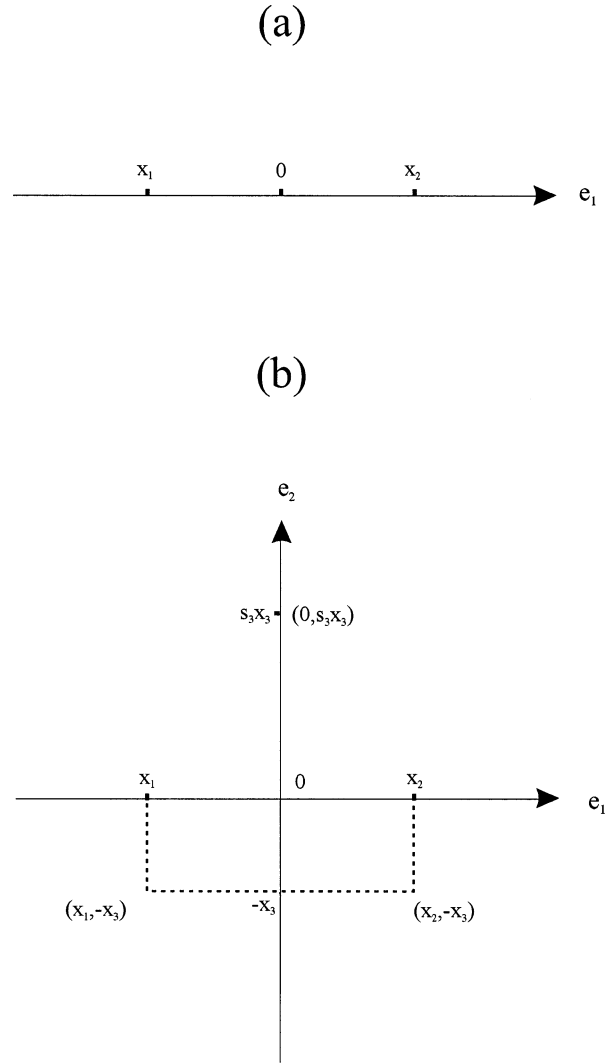


FIG. C1. Illustration of constructing spherical simplex points in (a) one dimension and (b) two dimensions.

So, we choose $s_3 = 2$ and $x_3 = 1/\sqrt{6}$. To satisfy the equal-distance constraint,

$$x_1^2 + x_3^2 = (s_3 x_3)^2. \quad (\text{C14})$$

By substituting x_1 , x_3 , and s_3 into (C14), it is easy to verify that (C14) is satisfied. So the three points in two dimensions are $(-1/\sqrt{2}, -1/\sqrt{6})^T$, $(1/\sqrt{2}, -1/\sqrt{6})^T$, and $(0, 2/\sqrt{6})^T$. Similarly, we can extend the set to three dimensions by extending the above three points to the direction \mathbf{e}_3 by $-x_4$ and a new point is added as $(0, 0, s_4 x_4)^T$. To satisfy the mean and covariance constraints in \mathbf{e}_3 , we choose $s_4 = 3$ and $x_4 = 1/\sqrt{12}$. This solution also satisfies the equal-distance constraint. Apply the same procedures to extend to the $(K - 1)$ dimensions. The solution \mathbf{U} in the format of (C8) is

$$\begin{bmatrix}
 \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \dots & \dots & 0 \\
 \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & 0 & 0 & \dots & \dots & 0 \\
 \frac{-1}{\sqrt{12}} & \frac{-1}{\sqrt{12}} & \frac{-1}{\sqrt{12}} & \frac{3}{\sqrt{12}} & 0 & \dots & \dots & 0 \\
 \vdots & \vdots & \vdots & \vdots & & & & \vdots \\
 \frac{-1}{\sqrt{j(j+1)}} & \dots & \dots & \dots & \frac{-1}{\sqrt{j(j+1)}} & \frac{j}{\sqrt{j(j+1)}} & 0 & 0 \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\
 \frac{-1}{\sqrt{(K-1)K}} & \dots & \dots & \dots & \dots & \dots & \frac{-1}{\sqrt{(K-1)K}} & \frac{K-1}{\sqrt{(K-1)K}}
 \end{bmatrix}. \quad (C15)$$

REFERENCES

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Buizza, R., and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–1456.
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145.
- Farrell, B. F., 1988: Optimal excitation of neutral Rossby waves. *J. Atmos. Sci.*, **45**, 163–172.
- , 1989: Optimal excitation of baroclinic waves. *J. Atmos. Sci.*, **46**, 1193–1206.
- Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560.
- Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential, and variational. *J. Meteor. Soc. Japan*, **75**, 181–189.
- Jeffery, T. K., J. H. Hack, B. B. Gordon, B. A. Boville, B. P. Briegleb, D. L. Williamson, and P. J. Rasch, 1996: Description of the NCAR community climate model (CCM3). NCAR Tech. Note NCAR/TN-420+STR, 152 pp.
- Julier, S. J., 2003: The spherical simplex unscented transformation. *Proc. IEEE American Control Conf.*, Denver, CO, IEEE, 2430–2434.
- , and J. K. Uhlmann, 2002a: Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations. *Proc. IEEE American Control Conf.*, Anchorage, AK, IEEE, 887–892.
- , and —, 2002b: The scaled unscented transformation. *Proc. IEEE American Control Conf.*, Anchorage, AK, IEEE, 4555–4559.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliaigis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- Nakos, G., and D. Joyner, 1998: *Linear Algebra with Applications*. Books/Cole, 666 pp.
- Ott, R. L., 1993: *An Introduction to Statistical Methods and Data Analysis*. 4th ed. Duxbury Press, 1051 pp.
- Ross, S., 1998: *A First Course in Probability*. Prentice Hall, 514 pp.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- , and —, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- Wilks, D. S., 2002: Smoothing forecast ensembles with fitted probability distribution. *Quart. J. Roy. Meteor. Soc.*, **128**, 2821–2836.